

Using a corpus-assisted discourse studies (CADS) approach to investigate constructions of identities in media reporting surrounding mega sport events: the case of the London Olympics 2012

Book or Report Section

Accepted Version

Jaworska, S. (2016) Using a corpus-assisted discourse studies (CADS) approach to investigate constructions of identities in media reporting surrounding mega sport events: the case of the London Olympics 2012. In: Lamond, I. R. and Platt, L. (eds.) Critical Events Studies: Approaches to Research. Leisure Studies in a Global Era. Palgrave Macmillan, Basingstoke, pp. 149-174. ISBN 9781137523846 doi: https://doi.org/10.1057/978-1-137-52386-0_8 Available at <http://centaur.reading.ac.uk/43261/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: http://dx.doi.org/10.1057/978-1-137-52386-0_8

Publisher: Palgrave Macmillan

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Using a corpus-assisted discourse studies (CADS) approach to investigate constructions of identities in media reporting surrounding mega sport events: the case of the London Olympics 2012

Sylvia Jaworska (University of Reading)

Mega sports events are a unique type of planned events in that they attract large audiences across the globe to experience ‘a joint “reality”’ (Billings and Hundley, 2010, p. 2). However, for most people, this experience is only possible through media. Through the choice of what and how to report, what is included and excluded, media largely control the dissemination and representations of such events. Discourse, understood here predominantly as language use, plays a crucial role in this process in that it can effectively shape public perceptions of (sport) events and the participating actors. For example, a heavy use of metaphors from the domain of war in reporting of international football matches constructs the events as armed conflicts between two nations fuelling nationalistic ambitions (Bishop and Jaworski, 2003). Celebrating a sportsman as a *British* champion when he wins, but as a *Jamaican-born* runner when the win is contested is another illustrative example demonstrating how media can frame and shift perspectives through specific language choices.¹ Such choices are never neutral. Commentators and journalists have a large pool of language items at their disposal, but when writing about a specific event, they may prefer some words or phrases over others, thus reflecting certain attitudes and ideological stances and propagating a particular version of events (Baker et al., 2013).

It needs to be kept in mind that the media versions of events are not solely a reflection of what media producers see as interesting, relevant and suitable. Media are driven by profit orientation and hence, what is produced also needs to appeal to the target audience. The audience is not just the immediate reader or listener, but often an imagined and more general one (Cotter, 2011). Thus, media will often draw on popular discourses assumed to be widely shared in order to respond to and attract more listeners and readers. Studying media discourses surrounding events can, therefore, offer considerable insights into how events are medially constructed and shed light on stances and ideologies that underpin such constructions and are widely circulated in society.

¹ This example refers to Linford Christie, whose win in the 1988 Olympic Games was contested on the grounds that he may have used performance enhancing drugs. He was eventually cleared of any wrongdoing.

Commonly media discourses are examined using approaches such as Discourse Analysis (DA) and especially its recent faction of Critical Discourse Analysis (CDA) (Fairclough, 1989; Wodak, 2001; van Dijk, 1993). Both approaches are qualitative in nature and are based on close readings of texts. By examining the interplay between context, text, and language, DA and CDA have been invaluable in providing insights into how media symbolically produce and reproduce norms and ideologies, especially of the powerful and dominant groups. However, both approaches operate normally with smaller sets of media data. While this can produce rich accounts of a particular issue shedding light on nuanced or subtle discursive strategies and representations, findings obtained from an analysis of a handful of articles are difficult to generalise beyond the studied cases. They may also tell us little about the more pervasive and widespread ways of how media construct events. As Fairclough (1989, p. 54) observes ‘the effects of media power are cumulative, working through the repetition of handling causality, particular ways of positioning the reader and so forth.’ It is for this reasons that analysts interested in media discourses are increasingly adopting the tools and methods of Corpus Linguistics (CL), which investigates linguistic patterns in large amounts of language data known as corpora. Corpora are normally interrogated using corpus-linguistic specialist software and quantitative, statistical techniques that allow for retrieval of frequencies, keywords and collocations. In this way, analysts can scan large amounts of data and reveal frequent and repeated linguistic patterns that are not immediately visible to the naked eye. Quantitative results produced by corpus tools are studied subsequently in more depth using qualitative discourse-analytic techniques. The combination of quantitative corpus techniques with more qualitative methods have led to fruitful methodological synergies (Baker et al., 2008; Baker and Levon, 2015) and approaches such as the corpus-assisted discourse studies (CADS) (Partington et al., 2013) that can reveal much more nuanced patterns of language use and representations in the media when compared with a quantitative or qualitative analysis alone. The main aim of this chapter is to demonstrate how the CADS approach can be effectively used to investigate media construction of events, in particular mega sport events. It will do so by takings as an example the media reporting surrounding the London Olympics 2012.

Sport, especially mega sports, events are an interesting site for social scientists and (critical) discourse analysts for many reasons. Sport is a major socialising force and a common cultural and symbolic resource deployed to aid solidarity, understanding and identity formation (Meân and Halone, 2010). Sport is also a site of tensions mostly due to ideological contradictions between its underlying values and the actual practice. Although

sport ethos is based on neutrality, inclusivity and fairness with the aim to promote international harmony, more often than not, it becomes a site of aggressive national rivalry and a strong reaffirmation of national pride. Despite some progress and changing social attitudes, it is still far away from being inclusive. Sport is recognised as one of the major masculine domains perpetuating the heterosexual norm (Daddario, 1994). Undoubtedly sport, due to its global impact, has the potential to be a vehicle of social transformations; there are numerous examples demonstrating the impact of sport on social and ethnic barriers (for example Carrington, 1998; Woodward 2004). At the same time, sport still continues to be an expression of privilege, domination and subordination maintained through the complex mechanisms of inclusion and exclusion (Jarvie, 1991; Long and Spracklen, 2011). It is this contradiction that makes sport ‘a disputed ideological terrain’ and a powerful site in the re/production of traditional hegemonic orders (Meân, 2010, p. 67). Thus, sport and especially its media representations are an important lens through which to examine social relations, and constructions of identities. Against this background, the paper focuses specifically on the investigation of discourses around identity that are constructed and supported by media coverage during major global sports events. The specific research questions that this study investigates are:

- 1) which identities are foregrounded and positively valued and which are backgrounded, negatively valued or absent?
- 2) do mega sports events have an impact on the discursive ways in which national, racial or gendered identities are constructed, and, if so, what is the nature of this effect?

The remainder of this chapter is structured as follows. Section 2 offers a brief overview of research concerned with media discourses surrounding sports events. Section 3 discusses in depth the CADS methodology, its principles and analytical tools. In Section 4, the main results are presented and discussed in light of the above research questions. The paper concludes with critical observations regarding the benefits and limitations of the adopted CADS approach to study media constructions of mega sport events.

2. Media Representations of Sport

Sport as a cultural and symbolic resource has been of interest to historians and sociologists for some time. Three major themes have been addressed in particular: 1) sport’s role in promoting nationalism (Kinkema and Harris, 1998; Rowe, 2003; Hogan, 2003; Smith and

Porter, 2004), 2) sport, ethnicity and racism (Jarvie, 1991; Long 2000; Long and Spracklen, 2011) and 3) sport as a signifier of hyper-masculinity and a site of gendered ideologies (Pirinen, 1997; Kane and Lenskyj, 1998; Jones et al., 1999; Wensig and Bruce, 2003; Duncan et al., 2005; Bruce et al., 2010). While it is beyond the scope of this chapter to discuss in detail this extensive body of research, some of the major conclusions need to be noted. Firstly, sport is seen as one of the major social forces that maintain a sense of national belonging. This role has become even more significant in the post-war world characterised by globalisation, mass migration and weaker nation states. As Smith and Porter (2004, p. 2) observe, increasingly, national identities are defined through and are ‘inextricably linked to what happens on the field of play.’ Secondly, sport continues to be a site of white privilege and domination (Spracklen, 2013). Although the participation of various ethnic groups in sport has considerably increased in the last decades leading some to celebrate sport as tool of equal opportunity, research shows that racism is still experienced at all levels of sport (Long and Spracklen, 2011). Arguably, overt expressions of racism seem to be less common, not least because of the many anti-racism campaigns introduced by various sporting bodies. However, racism still prevails, though in more subtle forms, for example, as institutionalised racism that restricts access of certain groups to positions of power. Although teams in some sports appear to be more diverse, we rarely find sportspeople from ethnic groups in the role of coaches, managers or members of sporting board rooms. Related to this is the issue of gender representation. Women participation in national and international tournaments has too increased significantly since 1950s, almost reaching parity with men in some domains, for example in Olympic sports. Nevertheless, research shows that sport continues to be a male domain that maintains traditional and heterosexual gendered hierarchies (Eastman and Billings, 2002; Duncan et al., 2005; Bruce et al., 2010; Meân 2010).

Methodologically, these issues have been mostly addressed by analysing historical records or using ethnographic and qualitative methods such as participant observations or interviews. Given the importance of media on experiencing sport events, some researchers have begun to look more closely at how media represent sport and sportspeople. There is now a large body of research documenting the media coverage of various sports events pointing to patterns of systematic mis- and underrepresentation. For example, Bruce et al. (2010) reveals that on average sportswomen receive less than 10 per cent of newspaper or TV coverage leading the authors to conclude that ‘women do not matter in our culture’ (Bruce et al., 2010, p. 5). Although the female converge in certain domains seems to increase (Pemberton et al., 2004), this does not necessarily lead to gender parity. As demonstrated by Duncan et al. (2005),

despite the increased coverage of women during Olympic Games, the reporting tends to depict female athletes in disciplines traditionally seen as feminine such as swimming, or gymnastics. Such disciplines focus heavily on aesthetics as opposed to strength and contact – attributes associated with male’s sports. Thus, some researchers are sceptical about the increased female coverage in media arguing that instead of achieving gender equality, it further perpetuates the ideology of biological difference which views women as suitable for certain sports only and denies them the possibility of demonstrating their prowess in others (Pirinen, 1997; Jones et al., 1999; Duncan et al., 2005, Bruce et al., 2010). This comes particularly into view in the ways in which female and male athletes are depicted. As Eastman and Billings (2002) show, sportswomen are more frequently described by references to their age, emotions, dating habits or family; these attributes that are rarely used to refer to sportsmen. Successful female athletes are also often compared to famous sportsmen. In contrast, successful male athletes are never compared to famous sportswomen, but to religious, mythical or fictional figures of power such as Jesus or Superman. Eastman and Billings (2002, p. 208) conclude that such comparisons place sportsmen several levels above sportswomen and install the men athlete as the prototypical athlete. Early research by Daddario (1994) also shows that sportswomen are likely to be represented as sex objects or in heterosexual roles pointing to the impact of sports media in perpetuating the norm of compulsory heterosexuality (Kane and Lenskyj, 1998).

Similar mis- and underrepresentation has been shown in research examining the construction of race in sports media. Billings and Eastman’s (2002) study of American telecasts of the 2000 Sydney Olympics demonstrate that white American men athletes were the most mentioned and most positively portrayed. Their success was associated with commitment, whereas black athletes were portrayed as having innate sporting abilities. Similarly, in a longitudinal content-analytic study on the representations of black athletes, Goss et al. (2010) demonstrates that certain features such as strength and hyper-aggressivity continue to be strongly associated with black players. Intelligence and leadership are, in contrast, more often attributed with white sportsmen (McDonald, 2010). Overall, sports media tend to over-signify blackness, whereas whiteness remains the unquestioned and ‘unmarked’ norm (Spracklen, 2013).

All in all, sports media have been shown to construct sport as a domain of hyper-masculine, white and heterosexual identities marked by aggressive and competitive attitudes (Meân, 2010). Furthermore, this hegemonic gender and racial representations seem to be strongly tied with national identities. Studies that examined the intersections of race, gender

and nationalism in the coverage of international sport events demonstrate the prevalence of nationalistic biases (Billings and Angelini, 2007) and the homogenising ideologies of nation states (Hogan, 2003). In fact, mega sports events are often directly utilised by governments to project the idea of a unified nation, while masking social inequalities and divides (Hogan, 2003). On the one hand, an increased coverage of women or members of ethnic groups serves this nationalising project in that it helps propagate the idea of a whole and ‘happy’ nation. On the other hand, media representations of sports events draw heavily on national discourses or in Hall’s (1992) sense, narratives of a nation, that is, images, stories, symbols, myths, metaphors and stereotypic ‘national’ characteristics that all together construct an experience of shared history and strengthen the sense of national unity. These are also adopted to create divides between nations, which further fuels nationalistic rivalry, as Bishop and Jaworski (2003) show in an analysis of media reporting surrounding the football match between Germany and England during Euro 2000. In this way, sports media perpetuate the homogenising and nationalistic ideologies and are significant contributors to the process of ‘nation imagining’ (Anderson, 1983).

Research studies concerned with media representations of sport discussed above highlight the role of sports media in constructing and perpetuating traditional, hegemonic hierarchies and ideologies. Due to various awareness-raising campaigns, overt expressions of racism or sexism are rare. Nevertheless, biases still prevail albeit they tend to be expressed in more subtle or ambivalent ways, for example, through the amount of attention given to and attributions associated with certain groups. While the conclusions from this research are compelling and consistent, there are some methodological issues. The dominant approach used is that of Content Analysis (CA). While CA has a number of strengths, it is based on coding schemes and categories that are normally set *ex ante*. The data is subsequently scanned for the absence or presence of the set categories or themes. While this may capture a number of relevant topics, some topics or categories may be omitted. Also, most of the studies are based on the analysis of a small amount of media data making generalisations difficult. Finally, many of the studies highlight the importance of discourse and language, but with a few exceptions (Bishop and Jaworski, 2003; Meân, 2010), most of them are not systematic analyses of discourse or linguistic choices. The corpus-assisted discourse studies (CADS) approach, which is a linguistic data-driven methodology, can help address some of these weaknesses and contribute to a more systematic understanding of the role of discourse in constructing events in the media. The next section moves on to outline the principles and key analytical tools of the CADS approach.

3. The Methodology of Corpus-Assisted Discourse Studies (CADS): principles and tools

The term CADS was first introduced by Partington (2004) in order to account for the growing body of research interested in discourse and adopting the techniques of Corpus Linguistics. Generally speaking, CADS is an offspring of the parent methodology of Corpus Linguistics (CL) and was much inspired by the early work by Stubbs (1995), Hardt-Mautner (1995) and Krishnamurthy (1996).² This section begins by describing the main concepts and tools used in corpus research. Subsequently, the specificity and tenets of CADS are discussed.

CL is primarily concerned with studying language on the basis of large electronic collections of real-life linguistic data normally known as corpora (McEnery and Hardie, 2012). Corpora can be quickly and reliably searched using specialist corpus-linguistic software programmes known as concordancers such as AntConc (Anthony, 2011), WordSmith Tools (Scott, 2008) and Sketch Engine (Kilgariff et al., 2004). Although the starting point for analysis will be a statistical output of some sort, for example a frequency list, corpora can also be interrogated qualitatively by studying selected outputs, for example, concordance lines. Generally speaking, CL is interested in studying quantitatively and qualitatively lexical and grammatical patterns across different varieties (for example, national varieties such as British English), domains (for example, business communication) or specific genre (business emails). Insights derived from corpus research have increased our understanding of language use by providing empirical evidence for the existence of regularities and patterns that are not immediately visible to the naked eye. As the father of CL John Sinclair once pointedly remarked: ‘The language looks rather different when you look at a lot of it at once’ (Sinclair, 1991, p. 100).

The key analytical tools underpinning most of corpus work are: frequency, keywords, collocation and concordance. These will now be discussed in more detail. In corpus linguistic terms, frequency refers to the count of items in a corpus, whereby *item* can be a word, a part of speech or a keyword. Frequency lists are a useful entry point to a corpus in that they can reveal items that occur often (on top of the list) or rarely (towards the bottom of the list). Frequent words can point to certain regularities, which, in turn, can be a sign of importance

² There has been a great deal of debate regarding the status of Corpus Linguistics. Some researchers see CL as a new theoretical approach to language, whereas others argue that CL is more of a methodological package (see Partington et al., 2013). The author of this chapter agrees with the view expressed by Partington et al. (2013, p. 7) that the use of corpora and corpus tools is not in itself a new theoretical advance in Linguistics, but it certainly is a new methodological paradigm that have the potential to revise some of the existing theories or concepts and lead to new ones.

and salience. Another good way of interrogating a whole corpus is via keywords. In Corpus Linguistics, a keyword is a word which occurs unusually frequent in a given corpus, as compared to another mostly larger reference corpus (Scott, 2010). The *keyness* of an item is usually established using a test of statistical significance, mostly log-likelihood. Whereas a frequency list indicates absolute or raw frequencies, a keyword list shows relative frequencies and it can be a useful tool in revealing the main themes of the studied corpus. Researchers often categories retrieved keywords manually into semantic groups in order to see the wider range of topics covered in a given corpus.

While frequency and keyword lists are useful in revealing frequent or unusually frequent items and can signpost major themes, they only show single lexical items and in isolation. We know from studies in phraseology that a single word does not necessarily encompass all the meanings that the word in question has. These often arise from the typical combinations of the word with other lexical items, that is, from collocations and from their use in context. Stubbs (2001, p. 105) gives an example of ‘cosy’, whose meaning generally tends to be positive, but in the collocation with ‘little relationship’, it expresses the negative meaning of ‘cliquey’. In Corpus Linguistics, collocation is understood as the co-occurrence of two or more words within a certain span, for example four items to the left and four to the right and a certain cut-off point (for example, occurring five times or more). A distinction is normally made between co-occurrences that are determined on the basis of raw frequency or significance testing (Barnbrook et al., 2013; McEnery and Hardie, 2012). The latter helps establish the strength of associations and commonly t-score, mutual information or log-likelihood are used for that purpose. It needs to be borne in mind that each test favours different types of words and hence yields different results (Baker, 2006, p. 100-4).

Collocations are very useful pointers of recurrent, typical lexical choices that are frequent and hence preferred or salient in a given data set. Such recurrent preferences are not just a matter of individual choices, but largely reflect established practices and evaluations and are often a means by which communities express, interpret and evaluate people and actions. For example, Stubbs (2002) shows how the word ‘gossip’ is mostly used with terms denoting or implying a female speaker highlighting the stereotype that only women gossip. Frequent patterns of co-occurring choices can, therefore, indicate how a phenomenon, group of people or events are persistently framed in discourse. As Stubbs (2001, p. 35) states, collocations are not simply lexical items, they ‘are also widely shared within a speech community’ and can act as ‘nodes around which ideological battles are fought’ (ibid., p. 188).

Finally, corpus researchers can study selected words or phrases in context by examining concordance lines. Concordances are lists with lines that display all occurrences of a search term. The word or phrase is normally positioned in the middle with a few words to the left and to the right. Such a display is known as a KWIC, short for Key Word in Context. Figure 1 provides an example of a random concordance for the plural noun ‘women’ from the COR-OLYMP corpus, which includes press articles published in major national British national newspapers during the London Olympics 2012 (see below). A close examination of concordance lines is a qualitative technique that enables the researcher to discover lexico-grammatical patterns offering clues to the uses and meanings of the search term in context. To illustrate an example, we will have a look at the concordance lines in Figure 1. As can be seen, the term ‘women’ often collocates with the modifier ‘British’, ‘Britain’s’ or ‘Great Britain’s’ highlighting potentially a national bias when referring to female athletes in the context of the Olympic Games. Another interesting pattern is the use of ‘women’ in the second position after ‘men’ providing some evidence for the male-firstness observed in general English usage (Baker, 2014). In this way, a concordance analysis can reveal associations that would otherwise be difficult to detect and could provide evidence for the existence of certain discourses and biases.

Figure 1 Concordance lines of ‘women’ in COR-OLYMP

N	Concordance
1	things. To see a packed Wembley stadium watching Britain's women beat Brazil was one such moment. How can I explain to my
2	,2.30pm) japanegypt(oldtrafford,noon) Great Britain's women produced another display that belonged in the category of
3	a shoulder and was also ruled out of the Games. Great Britain's women take on Belgium today still unsure if Kate Walsh, their captain,
4	in gymnastics in more than a century and medals won by men and women gymnasts on the world stage in the past decade have helped
5	in gymnastics in more than a century and medals won by men and women gymnasts on the world stage in the past decade have helped
6	moment. How can I explain to my son that football is for men and women when all he sees and reads is men? Well, he knows now. For
7	things. To see a packed Wembley stadium watching Britain's women beat Brazil was one such moment. How can I explain to my
8	really can become reality." Eight more of the 3,854 tall men and women who initially signed up are competing at London 2012 - another
9	members, I don't like to see that." More than 105,000 gay men and women have formalised their relationships, five times more than
10	grit and even genius; but not fanatical, drilled armies of sportsmen and women chosen for their big hands and feet. Where this country can
11	inexorable passage of time. And in purely national terms, the British women have been fab, medal-winners and also-rans and all. As I write,
12	have been fab, medal-winners and also-rans and all. As I write, British women have won 102/3 gold medals, 7Â? silvers and 6 bronzes.
13	rugby is known, and the ruthless goalball, in which men and women in blindfolds hurtle a heavy ball at each other at up to 60mph.
14	inexorable passage of time. And in purely national terms, the British women have been fab, medal-winners and also-rans and all. As I write,
15	have been fab, medal-winners and also-rans and all. As I write, British women have won 102/3 gold medals, 7Â? silvers and 6 bronzes.
16 Champion in Beijing, Nicole Cooke is one of four British women among the 67 riders taking part in the road race, which begins

The CADS approach utilises the quantitative tools offered by CL, but it extends the methodological paradigm by integrating techniques commonly associated with qualitative discourse analysis in order to understand the discourse in question and its context as much as possible. Alongside frequency or keyword lists, CADS researchers interrogate the data in a variety of ways, for example, by close reading, watching or listening to its subsets (Partington

et al., 2013, p. 12). In most cases, the studied phenomenon is further contextualised by considering and examining its social, political or historical context (Partington, 2014; Taylor, 2014). In contrast to general corpus research, CADS researchers prioritise a comparative approach. As Partington et al. (2013, p. 12) highlight: ‘we are not deontologically justified in making statements about the relevance of a phenomenon observed to occur in one discourse type unless, where it is possible, we compare how the phenomenon behaves elsewhere.’ Comparisons can take different forms. We can compare and contrast one discourse type in different sources, for example, national print newspapers vs. political speeches. We can also examine the same discourse produced under different circumstances, for example, during different events. Another type of comparison is a diachronic one, which involves studying a particular type of discourse at different points of time (Marchi, 2010; Taylor, 2010). This enables the researcher to detect changes and shifts over time. Finally, we can combine all the different types of comparison depending on the research questions.

The CADS approach has been successfully used to study a variety of discourses including discourse of science in the British press (Taylor, 2010), the discourse of morality also in the British press (Marchi, 2010), nationalism and language ideologies in the Canadian context (Vessey, 2013), the representations of Islam in the British media (Baker et al., 2013) and the portrait of immigrants in the British and Italian press (Taylor, 2014). Representations of specific events have not thus far been examined by using this methodology. Therefore, the next section presents a case study demonstrating how the CADS approach can be employed to investigate media constructions of events, especially the discourses of identity that underpin such constructions. The focus is on the London Olympics 2012.

4. Discursive constructions of identities in media reporting during global sports events: the case of the London Olympics 2012

This case study is based on the analysis of a large corpus of newspaper articles (COR-OLYMP) that were published in four major national British newspapers with their Sunday publications from 27 July to 12 August 2012 (see Table 1). The data was collected using the database Nexis. Instead of focusing solely on texts that were specifically about London Olympics, the decision was made to retrieve all articles published during this period. This allowed us to assess better the extent to which media construct sports events alongside other newsworthy events and which discourses are foregrounded and backgrounded in this context. Following the comparative CADS credo, we also decided to compare the media reporting

diachronically and created two control corpora encompassing articles from the same publications with the same time span, but one year BEFORE and one year AFTER the event. In doing so, we were able to evaluate the impact of mega sports events on the discursive ways in which such events and specifically identities during such events are medially constructed. Table 1 shows the data in all three corpora. Because of the unequal sizes of the data sets, results were normalised³ whenever necessary. The corpora were interrogated using the software programme WordSmith Tools Version 5 (Scott, 2008).

Table 1 Corpus Data

Corpus	BEFORE	COR- OLYMP	AFTER	Words total
Daily Mail/Mail on Sunday	1,135,264	1,114,492	1,087,458	3,337,214
Guardian/Observer	1,484,057	1,758,387	2,121,963	5,364,407
Times/Sunday Times	1,783,546	2,285,999	2,112,689	6,182,234
Sun	597,676	618,901	616,911	1,216,577
Total	5,000,543	5,777,779	5,322,110	16,100,432

As discussed in Section 3, frequency lists are a good point of entry into a corpus and thus, the analysis was started by examining and comparing the frequency lists retrieved from the three corpora. We included content words and personal pronouns only, as they are more likely to reveal aspects of identity constructions. Table 2 shows the most frequent content items in the three corpora.

Table 2 The 30 most frequent content words and pronouns in the three corpora

Rank	BEFORE	Freq.	NF ⁴	COR- OLYMP	Freq.	NF	AFTER	Freq.	NF
1.	he	33,354	667.0	I	39,572	684.9	he	38,610	725.5
2.	I	29,578	591.5	he	38,413	664.8	I	33,376	627.1
3.	his	26,763	535.2	his	30,403	526.2	his	31,208	586.4
4.	they	18,299	365.9	her	19,681	340.6	said	18,635	350.1
5.	you	16,178	323.5	you	19,235	332.9	you	18,310	344.0
6.	we	14,827	296.5	she	18,614	322.2	we	17,356	326.1

³ When comparing frequency results obtained from corpora of unequal sizes, it is important to normalise the raw frequencies to a common base, for example, per million words. Normalised frequencies (NF) are calculated using the following simple formula: $NF = (\text{number of examples of the word in the whole corpus} \div \text{size of corpus}) \times (\text{base of normalisation})$.

⁴ Here, the normalised frequencies (NF) were calculated using 100,000 as the base of normalisation. This base of normalisation was used throughout.

7.	her	12,119	242.4	said	18,265	316.1	their	17,032	320.0
8.	she	11,315	226.3	we	17,388	300.9	her	13,757	258.5
9.	year	9,753	195.0	their	16,171	279.9	she	13,057	245.3
10.	my	9,296	185.9	my	11,650	201.6	year	11,816	222.0
11.	time	7,865	157.3	year	10,716	185.5	my	9,874	185.5
12.	new	7,463	149.2	time	10,177	176.1	time	9,508	178.7
13.	people	7,212	144.2	years	8,267	143.1	new	9,244	173.7
14.	them	7,000	140.0	him	8,036	139.1	people	8,466	159.1
15.	him	6,864	137.3	them	7,852	135.9	them	8,319	156.3
16.	years	6,472	129.4	Olympic	7,759	134.3	years	8,048	151.2
17.	us	5,555	111.1	me	7,537	130.4	him	7,916	148.7
18.	your	5,483	109.6	world	7,515	130.1	me	6,371	119.7
19.	world	5,165	103.3	people	7,340	127.0	us	6,340	119.1
20.	our	5,161	103.2	new	7,334	126.9	our	6,142	115.4
21.	says	4,732	94.6	team	6,818	118.0	your	6,127	115.1
22.	UK	4,472	89.4	London	6,691	115.8	says	5,679	106.7
23.	day	4,462	89.2	gold	6,598	114.2	made	5,333	100.2
24.	week	4,338	86.8	your	6,576	113.8	make	5,263	98.9
25.	make	4,336	86.7	us	6,236	684.9	day	5,194	725.5
26.	way	4,331	86.6	our	5,831	664.8	way	5,169	627.1
27.	made	4,179	83.6	British	5,643	526.2	old	5,162	586.4
28.	old	4,138	82.8	day	5,324	340.6	world	4,911	350.1
29.	London	4,101	82.0	good	5,066	332.9	UK	4,885	344.0
30.	home	4,006	80.1	final	5,056	322.2	week	4,864	326.1

Table 2 points to a number of lexical items that denote national and gender identities and these seemed worthy of further scrutiny. We begin by examining the item ‘British’, which belongs to the 30 most frequent content items in COR-OLYPM and occurs much lower on the frequency list in the two other corpora. This could suggest that national identity is more emphasised during global sports events such as Olympic Games. To further test this claim, we compared the use of other typical items denoting national and regional identities used in the UK to see whether ‘British’ was preferred during the London Olympics.

Figure 1 *Terms denoting national and regional identity in the three corpora*

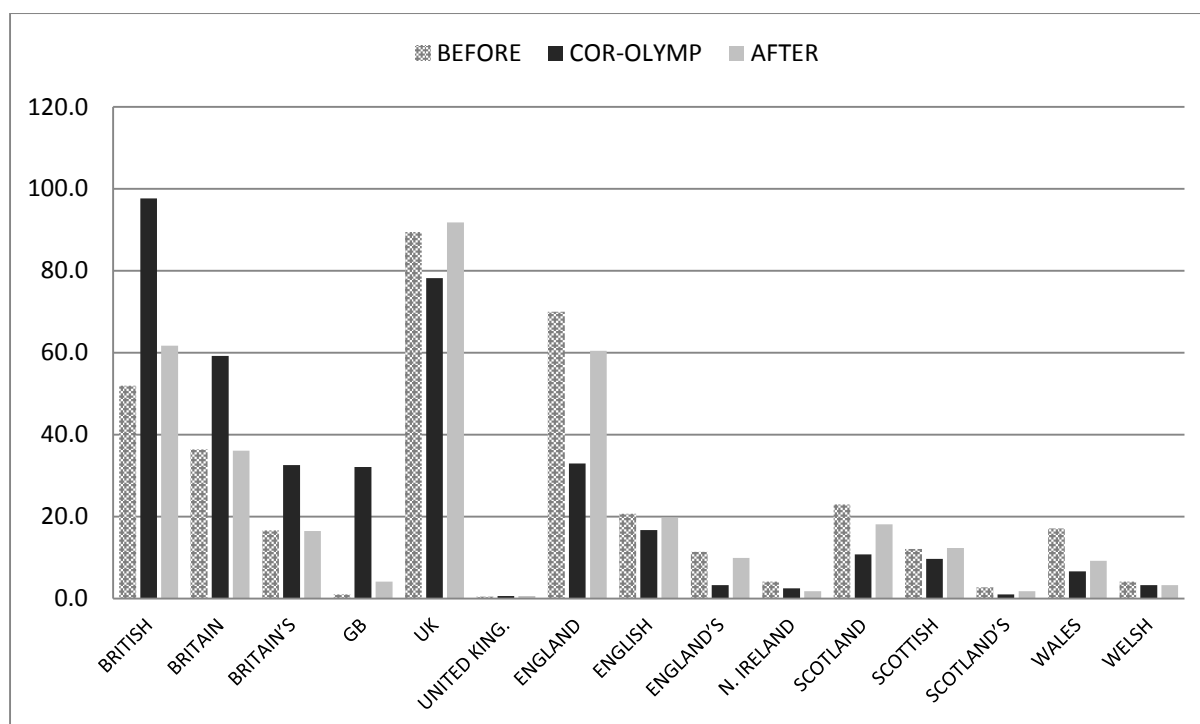
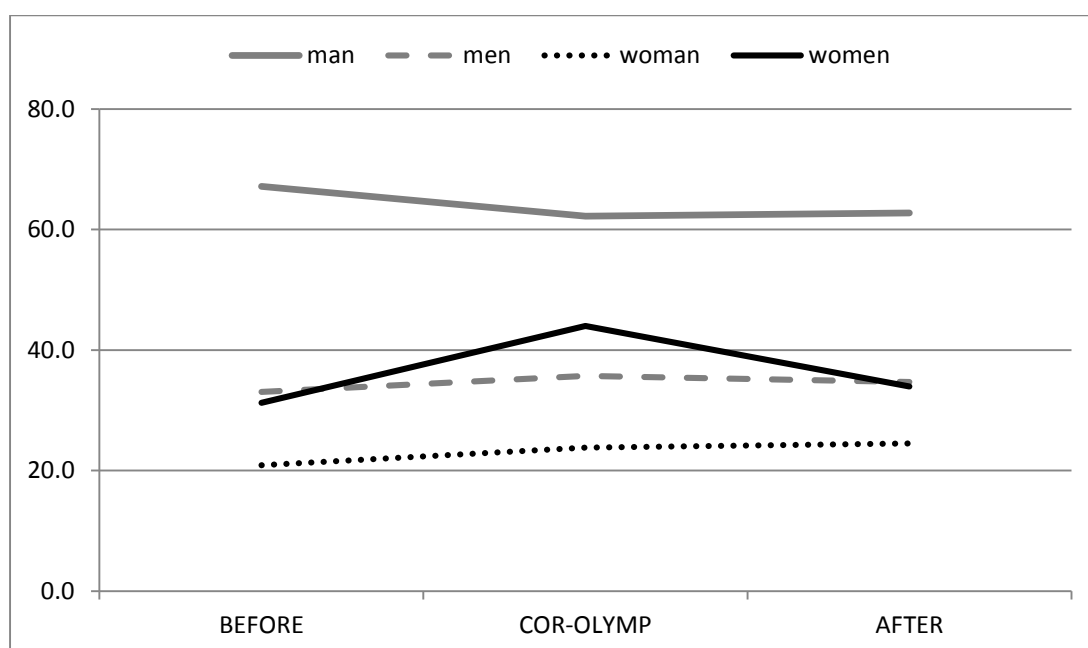


Figure 1 demonstrates a considerable increase in the use of ‘British’, ‘Britain’ and ‘Britain’s’ in COR-OLYMP. Together with ‘UK’ and ‘GB’, these were the preferred national terms. The frequent use of GB in the context of London Olympics is not surprising, as the acronym was used in the name of the British national team. Terms denoting regional identities such as ‘England’, ‘English’, ‘Scotland’ and ‘Scottish’ experienced a drop, in some cases substantial. Overall, the major national newspapers seem to promote the concept of nation-state or big-state nationalism more strongly than aspects of regional identity, as evidenced by the lesser attention given to Scotland, Wales and Northern Ireland in all three data sets. The London Olympics appears to have reinforced this trend. Thus, the data provides some empirical evidence suggesting that media reporting during global sport events effectively strengthen the concept of national identity and nation-state.

Other interesting items appearing on the frequency lists are words denoting gender. For example, the two masculine pronouns ‘he’ and ‘his’ belong to the most frequent items in all three corpora. It is notable that these pronouns always appear before the feminine pronouns ‘she’ and ‘her’ suggesting that in the media men receive more attention than women. This is consistent with previous corpus research that revealed a significantly higher proportion of male than female terms in general language usage in English (Sigley and Holmes, 2002; Pearce, 2008; Baker, 2014) and in the British media (Caldas-Coulthard and Moon, 2010). However, when we compare the normalised frequencies of the female pronouns across the

three subcorpora, an interesting result emerges. In the BEFORE Corpus, ‘she’ occurs 226.3 times per 100,000 words. In COR-OLYMP, the number rises to 322.2 and in the AFTER Corpus declines again to 245.3. The same pattern applies to the use of the pronoun ‘her’ as well as other terms denoting male and gender identities. Figure 2 shows the trends for ‘woman’, ‘women’, ‘man’ and ‘men’. Thus, although the traditional order was still maintained, the London Olympics seems to have had some impact on the gender representations in that media gave more attention to women as opposed to the time before and after the event.

Figure 2 *Distribution of gendered lexical items*



This makes gender an interesting area for further investigation. For reasons of space this cannot be perused here in greater depth. To illustrate the advantages of corpus tools and methods, this section focuses just on one example, that is, the collocations used with the term ‘women’. Table 3 demonstrates the strongest collocations as measured by t-score and retrieved within a -5 and +5 span. Only content words were included. As can be seen, items associated with ‘women’ vary across the corpora. In the BEFORE and AFTER corpus, women frequently co-occur with lexical words pointing to reproductive status (‘pregnant’, ‘children’), appearance (‘beautiful’, ‘hair’) and age (‘young’, ‘aged’, ‘older’). In the AFTER corpus, we also find references to crime (‘violence’, ‘rape’) and sex (‘sex’). Previous corpus research on gender representations in the media points to the same tendency: women seem to

be more strongly associated with appearance, sexuality and violence in contrast to men, who tend to be linked with careers, roles and status (Caldas-Coulthard and Moon, 2010; Pearce, 2008).

Table 3 *The 20 strongest collocations of ‘women’ across the three corpora*

BEFORE	COR-OLYMP	AFTER
men, their, young, children, said, aged, number, think, older, pregnant, look, make, girls, age, beautiful, middle, loose, time, want, our	men, young, first, Britain’s, British, said, she, can, sport, team, games, Olympic, final, make, her, gold, he, you, time, group	men, young, children, said, he, can, work, number, she, hair, rape, violence, pregnant, say, girls, sex, older, British, group, threats

Interestingly, however, is the absence of the ‘regular’ associations with women in the media reporting during the London Olympics. The collocations that appear in COR-OLYMP are, in some sense, ‘unusually’ positive and point to different discourses about women. Firstly, we find here two modifiers ‘Britain’s’ and ‘British’ suggesting that in the context of global sports events women are strongly attributed with the nation-state. Interestingly, the association between Britain and women appears predominantly in the context of celebrating sporting success. A quick look at concordance lines in Figure 3 illustrates this point: the attribution with ‘British’ and ‘Britain’ is clearly linked with winning, especially gold medals.

Figure 3 *Concordance lines of ‘British women’ and ‘Britain’s women’ on COR-OLYMP*

N Concordance
1 gave thanks to Heather Stanning and Helen Glover. The first **British women** ever to win Olympic gold in rowing had done so with
2 a shoulder and was also ruled out of the Games. Great **Britain's women** take on Belgium today still unsure if Kate Walsh, their
3 silver in the road race on Sunday, came in tenth. Great **Britain's women** produced another display that belonged in the category of
4 things. To see a packed Wembley stadium watching **Britain's women** beat Brazil was one such moment. How can I explain to my
5 FOR FIGHTING SPORTS Russia 18 medals WOMEN **Britain's women** athletes have won ten golds, putting them ahead of France,
6 passage of time. And in purely national terms, the **British women** have been fab, medal-winners and also-rans and all. As I
7 . Hockey's tale of two Kates ended with a victory hug as **Britain's women** secured their first Olympic medal in 20 years yesterday.
8 . Winning an Olympic medal is the headiest cocktail of all. **BRITISH women** have a distinguished pedigree in modern pentathlon. When
9 to row what Hosking later described as a "time trial". The **British women** had the race under control from a long way out, rowing with
10 men's 200m individual medley, qualifying in joint seventh. **Britain's women** freestylers put up a spirited show in the 4x200m relay final.
11 and English pairs partner Helen Glover became the first **British women** rowers ever to be crowned Olympic champs. They
12 with William Hill that this would be the first Games in which **British women** won more medals than the men. As it stands, with a week of
13 Lewis's bronze in the heptathlon was the only medal won by **British women**, the gender gap has been closing with each Games. But a
14 her bet and she has statistics on her side. There are more **British women** taking part in this Olympics than ever (262) and they make
15 Olympic legacy should be that the media do not forget that **British women** do sport and do it well. The Guardian could lead the pack

In the concordance lines, we find a number of positive associations highlighting physical strength (‘beat’ in line 4), persistence (‘secured their first Olympic medal’ in line 7) and power (‘had the race under control’ in line 9) – attributes that are more commonly associated with men and rarely with women (Pearce, 2008). Frequent references to ‘British’ and ‘Britain’s’ when describing women provide yet another evidence for the robustness of the concept of nationalism in the context of global sport events. The example also shows that

women are more likely to be positively represented, if they win the highest award, that is, a gold medal. The frequent collocations of ‘gold’ and ‘first’ (see table 3) provide further support for this claim. To test this assumption, we will now look briefly at the collocation profiles of two female athletes Jessica Ennis, who won a gold medal in the women’s heptathlon, and Rebecca Adlington, who received two bronze medals in swimming. For reasons of space, we focus on collocations surrounding the surnames of the athletes only.

The first fact to note is that ‘Ennis’ (as the surname of Jessica Ennis) appears 745 times as opposed to ‘Adlington’, which occurs 384 times. This suggests that Jessica Ennis as a gold winner received almost double amount of attention in the media than the double bronze medallist Rebecca Adlington. Table 4 shows the 30 strongest associations with both terms.

Table 4 *Collocations of ‘Ennis’ and ‘Adlington’ in COR-OLYMP*

ENNIS	ADLINGTON
Heptathlon, best, Jessica, Farah, Wiggins, Rutherford, Mo, GB, hurdles, Bradley, poster, Olympic, stadium, stormed, medal, jump, gold, Sheffield, champion, winning, Britain’s, girl, star, greatest, refused, record, final, pressure, British, team	Rebecca, freestyle, swimmer, swum, bronzes, swimmers, Beijing, medals, bronze, Olympic, swimming, pool, women’s, won, champion, crowd, fifth, finished, win, double, race, herself, Miss, behind, goes, final, took, set, world, her

A number of interesting patterns emerge from the lists. Jessica Ennis is mentioned together with other male gold medallists, whereas Rebecca Adlington is not. Ennis is also described in superlatives as ‘best’ or ‘greatest’, which reflects her status as a champion. We also find references to ‘Britain’s’ and ‘British’ suggesting that her sporting success was framed in the national terms - Ennis as incorporation of the nation’s success. Interestingly, the two modifiers do not appear in the vicinity of ‘Adlington’ implying that bronze medallists do not ‘enjoy’ the same status. The collocations occurring with ‘Adlington’ point to her sport discipline and we do not find here any instances of positive associations. Interesting is the grammatical collocation ‘herself’, which occurs five times in the vicinity of ‘Adlington’. As a reflexive pronoun, herself ‘reflects back’ on the subject and is the object of the verb to which the action of the verb is ‘done’. It can also be used as an intensifier. Studying the context in which ‘herself’ collocates with ‘Adlington’ points to a rather negative discourse expressing either an apologetic stance and desperation, as the examples of expanded concordances below show:

- (1) By now people were starting to panic, not least Adlington herself. Sensing that the situation demanded it, she tried to make a move. But she just could not summon up the change of pace she needed.
- (2) But she cruised past Adlington, who then found herself trying to hang on to bronze.
- (3) But look just a little harder and you will find that Adlington, as she said herself, has ‘swum faster than this all year’. ‘I would have liked the time to have been a bit quicker. I’m not going to lie,’ Adlington said. ‘I’ve done that time all year and I don’t know what happened. Everything kind of caught up with me.’

Although the collocation profiles of the two athletes differ considerably, there is some similarity. Striking is the use of the term ‘girl’ as a collocation of ‘Ennis’ and ‘Miss’ in references to ‘Adlington’. Referring to an adult woman as a girl can be seen as patronising and trivialising. Previous corpus research has noted the tendency of the terms ‘woman’ and ‘girl’ being used synonymously, whereas this does not seem to apply to ‘man’ and ‘boy’ (Taylor, 2013). The representation above echoes this tendency and supports a claim suggested by the linguist Dwight Bolinger (1980, p. 100) that ‘a female never grows up.’ Also, the use of ‘Miss’ in references to Rebecca Adlington shows that women are likely to be described by pointing explicitly to their marital status – something which has also been identified as a common feature in descriptions of women in the media.

The above analysis has revealed a number of notable discursive patterns and tendencies of media reporting during the London Olympics. The frequent use of items such as ‘British’ and ‘Britain’ and decline in regional terms exemplify how global sport events are used to boost national unity and the concept of nation-state. The persistent association with ‘British’ and ‘women’ provides further evidence for the dominance of national bias, which frames female success predominantly as a symbol of ‘national’ success. In terms of gendered discourses, we can note a pattern of ‘unusually’ positive representations of women achieving almost parity with men. However, as the analysis of collocations profiles of Jessica Ennis and Rebecca Adlington exemplary shows, only those women who achieve the highest award are likely to be positively represented. The analysis also demonstrates that despite the generally positive and less gendered representations, the media still tend to adhere to certain trivialising and patronising tendencies when referring to adult women. The use of ‘girl’ or ‘Miss’ illustrate this pattern.

Overall, the results seem to point to salient discursive patterns regarding identity constructions during a global sport event. However, the analysis focused on a single case and

limited number of examples. Thus, the identified patterns need to be verified by comparing, for example, female with male representations or representations of sportspeople of various ethnic backgrounds. Also, to see whether such discourses are typical for global sport events, a comparison with other sports events would be necessary.

5. Conclusions

The above analysis serves as a case study to illustrate how the CADS approach can be effectively used to study media constructions of global sport events. By scrutinising frequency lists and studying collocation profiles of selected lexical items denoting national and gender identity, we were able to identify prominent discourses around identity that the British press foregrounded during the London Olympics 2012. The comparative CADS approach has also allowed us to assess the saliency of such discourses: comparing the media constructions of the London Olympics with media reporting during the same time but one year before and one year after has demonstrated the prominence of nationalistic bias and positive female representations, though the latter appeared to serve more nationalistic tendencies rather than gender equality (Wensig and Bruce, 2003). For reasons of space, the analysis focused on selected analytical tools and procedures. Issues involved in corpus creation, data sampling and the use of statistical tests were not considered and readers are encouraged to consult, for example, McEnery and Hardie (2012) and Baker (2006) to find out more about these important aspects.

All in all, one of the major benefits of the CADS approach is its quantitative basis. Examining large amounts of textual data enables the researchers to produce empirical results that are more generalisable than those obtained from studying a few texts. It also fosters a greater distance to the data and increases objectivity of research in that it can help reduce some of the cognitive biases such as the primacy effect or confirmation bias (Baker, 2006, p. 10-12). Moreover, corpus analysis enables us to see frequent and repeated patterns that may not be immediately visible to the naked eye perusal or simply run counter to our intuition leaving room for serendipitous effects (Partington, 2014). Repeated associations are important because they help reveal discourses that are continually and systematically disseminated, thus, increasing our understanding of the recurrent or dominant ways of talking. It is precisely the ability of a corpus approach that it allows us 'to see which choices are privileged, giving evidence for mainstream, popular or entrenched ways of thinking.'

(Baker et al., 2013, p. 25). Also, by combining the quantitative corpus tools with qualitative discourse-analytical techniques, the CADS approach permits us to study our data comprehensively. By retrieving frequencies or keywords, we can identify general salient patterns in our corpus. These can be subsequently examined in more depth using qualitative techniques, for example, close readings of selected stretches of discourse via concordance lines. In this way, we may discover much more subtle and nuanced meanings that a list of keywords or collocations cannot reveal.

Despite its numerous benefits, a corpus methodology such as CADS has some limitations, especially when adopting the approach to study events and their mediatisation. Firstly, CADS is mostly concerned with texts and words, and leaves out the visual type of data. This could be a problem when examining media constructions of events, for which the visual component is central. A combination of CADS with a multimodal analysis would benefit future corpus-based research, though arguably this presents methodological and analytical challenges. Also, CADS is primarily interested in studying press data. Given the growing importance of social media in disseminating events of all kinds, future research needs to give more attention to the online mediatisation of events. Finally, corpus based analyses of media often assume that frequent discourses or representations are influential and widely shared in society. While this may be true, corpus-based results are rarely tested against views and attitudes of the wider public. The issue of reception and production of media contents is indeed rather neglected. Thus, another useful direction for future corpus-based research would be integration and triangulation of corpus methods with other qualitative and quantitative techniques used in social sciences, for example, ethnography or surveys. This would allow the researchers to validate some of the insights derived from corpus-based analyses, which, in turn, could offer invaluable insights into how discourses that underlie media representations of events really ‘work’ in a given discourse community.

6. References

- B. Anderson (1983) *Imagined Communities: Reflections on the Origin and Spread of Nationalism* (London: Verso).
- L. Anthony (2011) *AntConc*. Version 3.2.2 (Tokyo: Waseda University),
<http://www.antlab.sci.waseda.ac.jp/>
- P. Baker (2006) *Using Corpora in Discourse Analysis* (London: Continuum).
- P. Baker (2014) *Using Corpora to Analyse Gender* (London: Bloomsbury).

- P. Baker and E. Levon (2015) 'Picking the right cherries? A comparison of corpus-based and qualitative analyses of news articles about masculinity', *Discourse & Communication*, 9(2), 221–236.
- P. Baker, C. Gabrielatos, M. KhosraviNik, M. Krzyzanowski, T. McEnery and R. Wodak (2008) 'A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press', *Discourse & Society*, 19(3), 273–306.
- P. Baker, C. Gabrielatos and T. McEnery (2013) *Discourse Analysis and Media Attitudes* (Cambridge: Cambridge University Press).
- G. Barnbrook, O. Mason and R. Krishnamurthy (2013) *Collocation. Applications and Implications* (London: Palgrave Macmillan).
- A. Billings and J. Angelini (2007) 'Packaging the games for viewer consumption: Gender, ethnicity, and nationality in NBC's coverage of the 2004 Summer Olympics', *Communication Quarterly*, 55(1), 95–111.
- A. Billings and S. T. Eastman (2002) 'Selective representation of gender, ethnicity, and nationality in American television coverage of the 2000 Summer Olympics', *International Review for the Sociology of Sport* 37, 351–370.
- A. Billings and H. Hundley (2010) 'Examining identity in sports media', in H. Hundley and A. Billings (eds.) *Examining Identity in Sports Media* (Thousand Oaks, CA: Sage), p. 1–15.
- H. Bishop and A. Jaworski (2003) "'We Beat 'em': Nationalism and the hegemony of homogeneity in the British press reportage of Germany versus England during Euro 2000", *Discourse & Society*, 14(3), 243–271.
- D. Bolinger (1980) *Language, the Loaded Weapon* (London: Longman).
- T. Bruce, J. Hovden and P. Markula (eds.) (2010) *Sportswomen at the Olympics: A global comparison of newspaper coverage* (Rotterdam, The Netherlands: Sense).
- C. Caldas-Coulthard and R. Moon (2010) 'Curvy, hunky, kinky: using corpora as tools for critical analysis', *Discourse & Society*, 21(2), 99–133.
- B. Carrington (1998) 'Sport, Masculinity and Black Cultural Resistance', *Journal of Sport and Social Issues*, 22, 275–298.
- C. Cotter (2010) *News Talk: Investigating the Language of Journalism* (Cambridge: Cambridge University Press).
- G. Daddario (1994) 'Chilly scenes of the 1992 winter games: the mass media and the marginalization of female athletes', *Sociology of Sport Journal*, 11(3), 275–88.

- M. C. Duncan, M. A. Messner and N. Willms (2005) *Gender in televised sports: News and highlights shows, 1989-2004* (Los Angeles, CA: Amateur Athletic Foundation of Los Angeles), http://www.aafla.org/11pub/over_frmst.htm
- S. T. Eastman and A. Billings (2000) 'Sportscasting and sports reporting. The Power of gender bias', *Journal of Sport & Social Issues* 24(2), 192–213.
- N. Fairclough (1989) *Language and Power* (London: Longman).
- B. Goss, A. Tyler and A. Billings (2010) 'A Content Analysis of Racial Representations of NBA Athletes on Sports Illustrated Magazine Covers, 1970-2003' in H. Hundley and A. Billings (eds.) *Examining Identity in Sports Media* (Thousand Oaks, CA: Sage), p. 173–194.
- S. Hall (1992) 'The question of cultural identity' in S. Hall, D. Held and T. McGrew (eds.) *Modernity and its futures* (Cambridge, MA: Polity), p. 273–326.
- G. Hardt-Mautner (1995) 'Only connect: critical discourse analysis and corpus linguistic', *UCREL Technical Paper* 6 (Lancaster: University of Lancaster), http://ucrel.lancs.ac.uk/tech_papers.html
- J. Hogan (2003) 'Staging the nation: gendered and ethnicized discourses of national identity in Olympic opening ceremonies', *Journal of Sport and Social Issues*, 27 (2), 100–123.
- G. Jarvie (ed.) (1991) *Sport, Racism and Ethnicity* (London: Routledge).
- R. Jones, A. J. Murrell and J. Jackson (1999) 'Pretty versus powerful in the sports pages: print media coverage of us women's Olympic gold medal winning teams', *Journal of Sport and Social Issues* 23(3), 183–92.
- M. J. Kane and H. Lenskyj (1998) 'Media treatment of female athletes: Issues of gender and sexualities' in L. A. Wenner (ed.) *MediaSport* (London: Routledge), p. 186–201.
- A. Kilgarriff, P. Rychlý, P. Smrz and D. Tugwell (2004) 'The Sketch Engine', in *Proc EURALEX 200* (Lorient: France), p. 105–116.
- K. M. Kinkema and J. C. Harris (1998) 'MediaSport studies: key research and emerging issues' in L. A. Wenner (ed.) *MediaSport* (London: Routledge), p. 27–54
- R. Krishnamurthy (1996) 'Ethnic, racial and tribal: The language of racism?' in C. Caldas Coulthard and M. Coulthard (eds.) *Texts and Practices: Readings in Critical Discourse Analysis* (London: Routledge), p. 129–149.
- J. Long (2000) 'No racism here? A preliminary examination of sporting innocence', *Managing Leisure*, 5(3), 121–133.
- J. Long and K. Spracklen (2011) 'Positioning Anti-Racism in Sport and Sport in Anti-Racism' in J. Long and K. Spracklen (eds.) *Sport and Challenges to Racism* (Basingstoke: Palgrave), p. 3– 18.

- A. Marchi (2010) 'The moral in the story: a diachronic investigation of lexicalised morality in the UK press', *Corpora*, 5(2), 161–189.
- M. G. McDonald (2010) 'The Whiteness of Sport Media/Scholarship' in H. Hundley and A. Billings (eds.) *Examining Identity in Sports Media* (Thousand Oaks, CA: Sage), p. 153–172.
- T. McEnery and A. Hardie (2012) *Corpus Linguistics. Method, Theory and Practice*. (Cambridge: Cambridge University Press).
- L. Meân (2010) 'Making masculinity and framing femininity: Fifa, soccer, and world cup web sites' in H. Hundley and A. Billings (eds.) *Examining Identity in Sports Media* (Thousand Oaks, CA: Sage), p. 65–86.
- L. Meân and K. Halone (2010) 'Sport, language and culture: issues and intersections', *Journal of Social Psychology*, 29 (3), 253–260.
- A. Partington (2004) 'Corpora and discourse, a most congruous beast', in A. Partington, J. Morley and L. Haarman (eds.) *Corpora and Discourse* (Frankfurt/M: Peter Lang), p. 9–18.
- A. Partington (2014) 'Mind the gaps. The role of corpus linguistics in researching absences', *International Journal of Corpus Linguistics*, 19(1), 118–146.
- A. Partington, A. Duguid and C. Taylor (2013) *Patterns and Meanings in Discourse. Theory and Practice in Corpus-Assisted Discourse Studies (CADS)* (Amsterdam: Benjamins).
- M. Pearce (2008) 'Investigating the collocational behaviour of MAN and WOMAN in the BNC using Sketch Engine', *Corpora* 3 (1), 1–29
- R. M. Pirinen (1997) 'The construction of women's positions in sport: a textual analysis of articles on female athletes in finish women's magazines', *Sociology of Sport Journal*, 14(3), 283–301.
- M. Scott (2008) *WordSmith tools. Version 5*. Lexical Analysis Software.
<http://www.lexically.net/wordsmith/version5/index.html>
- M. Scott (2010) 'Problems in investigating keyness, or clearing the undergrowth and marking out trails', in M. Bondi and M. Scott (eds.) *Keyness in Texts* (Amsterdam: Benjamins), p. 43–58.
- R. Sigley and J. Holmes (2002) 'Looking at girls in Corpora of English', *Journal of English Linguistics*, 30(2), 138–157.
- J. Sinclair (1991) *Corpus, Concordance, Collocation* (Oxford: Oxford University Press).
- A. Smith and D. Porter (2004) *Sport and National Identity in the Post-War World* (London and New York: Routledge)
- K. Spracklen (2013) *Whiteness and Leisure* (Basingstoke: Palgrave).

- M. Stubbs (1995) 'Collocations and semantic profiles: on the cause of the trouble with quantitative studies', *Functions of Language*, 2(1), 23–55.
- M. Stubbs (2001) *Words and Phrases. Studies in Lexical Semantics* (London: Blackwell).
- C. Taylor (2010) 'Science in the news: a diachronic perspective', *Corpora* 5(2), 221–250.
- C. Taylor (2013) 'Searching for similarity using corpus-assisted discourse studies', *Corpora*, 8 (1), 81–113
- C. Taylor (2014) 'Investigating the representation of migrants in the UK and Italian press: A cross-linguistic corpus-assisted discourse analysis', *International Journal of Corpus Linguistics*, 19(3), 368–400.
- T. van Dijk (1993) *Elite Discourse and Racism* (London: Sage).
- R. Vessey (2013) 'Challenges in cross-linguistic corpus-assisted discourse studies', *Corpora*, 8(1), 1–26.
- E. Wensig and T. Bruce (2003) 'Bending the rules: media representations of gender during an international sporting event', *International Review for the Sociology of Sport*, 38, 387–396.
- R. Wodak (2001) 'What CDA is about - summary of its history, important concepts and its developments' in R Wodak and M. Meyer (eds.) *Methods of Critical Discourse Analysis* (London: Sage), p. 1–13.
- K. Woodward (2004) 'Rumbles in the Jungle: Racialisation and the Performance of Masculinity', *Leisure Studies*, 23, 5–17.